

This document is published in:

*Journal of Financial Econometrics*, 12, 1, 151-173, 2014

DOI:<http://dx.doi.org/10.1093/jjfinec/nbt011>

© Oxford University Press

# On the Properties of Regression Tests of Stock Return Predictability Using Dividend-Price Ratios\*

Seongman Moon <sup>†</sup>

Universidad Carlos III de Madrid

Carlos Velasco <sup>‡</sup>

Universidad Carlos III de Madrid

February 12, 2013

## Abstract

This paper investigates, both in finite samples and asymptotically, statistical inference on predictive regressions where time series are generated by present value models of stock prices. We show that regression-based tests, including robust tests such as the conditional test and the  $Q$ -test, are inconsistent and thus suffer from lack of power in local-to-unity models for the regressor persistence. The main reason is that, despite the near-integrated dividend-price ratio, the convergence rates of the estimates are slowed down because the present value model implies a shrinking innovation variance on the predictor, an effect which is masked in a predictive regression analysis with exogenous constant covariance of innovations. We illustrate these properties in a simulation study.

JEL Classification: C12, C22, G1.

Keywords: present value model, predictive regression, local-to-unity assumption, conditional test,  $Q$ -test,  $t$ -test.

---

\*This is a revised version of the paper circulated under the title "On the Properties of Regression Tests of Asset Return Predictability". We are grateful to Associate Editor and two referees for comments that lead to a much improved version of the paper.

<sup>†</sup>Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe Madrid, SPAIN. Email: smoon@eco.uc3m.es, Tel.: +34-91-624-8668, Fax: +34-91-624-9329. Research support from Spanish Secretary of Education (SEJ2011-0001) is gratefully acknowledged.

<sup>‡</sup>Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe Madrid, SPAIN. Email: Carlos.Velasco@uc3m.es, Tel.: +34-91-624-9646, Fax: +34-91-624-9329. Research support from the Spanish Plan Nacional de I+D+I (SEJ2007-62908 and ECO2012-31748) is gratefully acknowledged

## 1 Introduction

Predictable excess returns have been persistently documented in various asset markets using finance variables as predictors, for example, dividend-price ratios in stock markets and forward discounts in foreign exchange markets. However, some doubt on the validity of the typical econometric methods used in those studies has been casted. As Mankiw and Shapiro (1986) illustrate, the main concern is that the  $t$ -test may reject too often based on conventional critical values when the regressor is strongly persistent and its contemporaneous innovations are highly correlated with the returns. Several studies [e.g., Jansson and Moreira (JM, 2006), Campbell and Yogo (CY, 2006), and references therein] have developed robust tests to overcome this problem. Following Phillips (1987), these studies typically assume that the regressor follows a local-to-unit-root process in the predictive regression and the covariance matrix of innovations to asset returns and to predictors is parameterized independently of the regressor persistence process.

The present paper deviates from this line of research and let economic models determine the innovations covariance matrix.<sup>1</sup> We follow this path because those assumptions on persistency and innovations distribution are not independent in economic models. In particular, we show that the covariance of innovations is a function of the regressor persistence parameter in the typical present value model of stock prices. In turn, the relative variance of the two innovations is also a function of this parameter, significantly affecting statistical inference. For example, it is well known that the estimated OLS slope coefficient converges at a fast  $T$  rate in the predictive regression with the local-to-unity assumption rather than the usual  $\sqrt{T}$  rate. At the same time, however, this assumption also affects the innovations covariance matrix in present value models and thus the convergence rate of the estimates. This effect on the convergence rate is masked in the analysis of predictive regressions with an exogenous covariance matrix.

We analytically show that regression-based tests, including both JM's conditional test and CY's  $Q$ -test, are not consistent and thus suffer from lack of power for testing predictability of excess stock returns if the series are generated from present value models. This results because the faster convergence rate due to (near) non-

---

<sup>1</sup>This problem was first observed by West (2008) who shows that the  $t$ -test is not consistent if the discount factor in the present value model tends to be one with sample size.

stationarity is offset by the local-to-unity persistence effect on the relative variance of the two innovations in the predictive model. Monte Carlo experiments confirm our analysis in that the regression-based tests have power deficiency under a class of  $T^{-1}$  local alternatives despite of the nonstationary character of the regression.

In terms of exploiting restrictions in the predictive regressions implied by economic theory or accounting identities, our paper is related to recent studies such as Campbell and Thompson (2008), Cochrane (2008), and West (2008). For example, Campbell and Thompson (2008) impose restrictions on the signs of coefficients and return forecasts implied by investment theory, and show that those restrictions improve the out-of sample performance of key forecasting variables; Based on the approximate identity from the Campbell and Shiller (1988b) linearization of the definition of stock return, Cochrane (2008) illustrates that evaluating the joint distribution of the dividend-price ratio coefficients in the return and dividend growth regressions gives more powerful tests of predictability than doing it in each regression in isolation; West (2008) introduce a local-to-unity assumption on the discount factor in the present value model of foreign exchange rates to exploit the dependency between an economic parameter and the innovation covariance matrix in the foreign excess return predictive regression.

The organization of the paper follows. Section 2 presents the present value model of stock prices for the predictive regressions and Section 3 shows the asymptotic properties of the regression-based tests under the assumption of local-to-unity persistence in the regressor and the covariance matrix of innovations implied by the present value model. Section 4 provides simulation results of those regression-based tests and conclusions follow.

## 2 Predictive Regressions

### 2.1 The Standard Predictive Regression Model

Consider the typical bivariate regression model with observations,  $t = 1, \dots, T$ ,

$$\begin{aligned} y_t &= \alpha + \beta x_{t-1} + u_t \\ x_t &= \gamma + \phi x_{t-1} + v_t, \end{aligned} \tag{1}$$

where  $u_t$  is a prediction error and  $x_0$  is fixed and known. For example,  $y_t$  is an excess stock return on a riskless interest rate,  $x_{t-1}$  is a predictor such as a dividend-price ratio, and the hypothesis of interest is  $\beta = \beta_0 = 0$  in the regression for testing

the predictability of excess stock returns [see, for example, Campbell and Shiller (1988a), Fama and French (1988), CY and references therein].<sup>2</sup>

The parameter  $\phi$  measures the degree of persistence in  $x_t$ : if  $\phi = 1$  then  $x_t$  is integrated of order one; if  $|\phi| < 1$  then  $x_t$  is integrated of order zero. The covariance matrix of the error terms,  $u_t$  and  $v_t$ , in the regression is denoted as

$$\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}. \quad (2)$$

As is well known, the Gauss-Markov theorem does not apply in this regression if  $\phi$  is equal to one (or not constant) and the contemporaneous covariance between error terms,  $\sigma_{uv}$ , is not zero. Most studies on statistical distortions in the predictive regression are mainly concerned with the effects of strong persistency on the shape of the distribution of the OLS estimate,  $\hat{\beta}$ , while typically assuming that  $(u_t, v_t)'$  are independently distributed  $N(0, \Sigma)$  with an exogenously given constant covariance  $\Sigma$ . However, as shown in the subsequent sections, the assumption of the constant covariance is not compatible with that of the local-to-unity regressor in the typical economic models of asset prices, since the covariance matrix elements depend on primitive parameters.

## 2.2 A Predictive Regression in a Present Value Model of Stock Prices

Present value models of asset prices have been widely used in stock markets as well as in foreign exchange markets. For example, Campbell and Shiller (1987, 1988b) use them for studying the behavior of stock prices and the term structure of interest rates, while Engel and West (2005) use them to study the link between fundamentals and exchange rates. We derive the covariance matrix from these well-known models in order to study how this implied specification by the present value models affects statistical inference. In particular, we focus on the present value model of stock prices which relates current prices to future dividends and returns.

Define a gross return,  $R_{t+1}$ , on the stock held between time  $t$  and  $t + 1$  by

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t},$$

---

<sup>2</sup>Or,  $y_t$  is a foreign excess return,  $x_{t-1}$  is a forward discount, and the hypothesis of interest is  $\beta = 0$  in the regression for testing uncovered interest parity (UIP) [see, for example, Lewis (1995) and Engel (1996) for a survey of such tests].

where  $P_t$  is an ex dividend-price and  $D_{t+1}$  is the dividend per share at  $t + 1$ . This equation can be rewritten by

$$\frac{P_t}{D_t} = R_{t+1}^{-1} \left( 1 + \frac{P_{t+1}}{D_{t+1}} \right) \frac{D_{t+1}}{D_t}, \quad (3)$$

and, in logs

$$p_t - d_t = -r_{t+1} + \ln(1 + \exp(p_{t+1} - d_{t+1})) + \Delta d_{t+1}, \quad (4)$$

where lower case letters denote logs of variables. Since a term  $\ln(1 + \exp(p_{t+1} - d_{t+1}))$  in the right-hand side of equation (4) is nonlinear, following Campbell and Shiller (1988b), we take log-linear approximation around a point,  $\exp(\overline{p - d}) = \overline{P/D}$ ,

$$d_t - p_t = r_{t+1} - \kappa + b(d_{t+1} - p_{t+1}) - \Delta d_{t+1}, \quad (5)$$

where  $b = \exp(\overline{p - d}) / (1 + \exp(\overline{p - d})) = \exp(g - r)$  and  $\kappa = -(1 - b) \ln(1 - b) - b \ln b$ . The second equality in the definition of the discount factor  $b$  is derived from equation (3) where  $g = \ln G$  is the constant dividend growth rate and  $r = \ln R$  is the constant log gross return. Here, we took the log linearization on the return identity by assuming that all the relevant variables are stable. Below we will provide our justification.

Taking conditional expectation on return identity (5), iterating forward it, and assuming that no rational bubbles exist,<sup>3</sup> we have the present value identity,

$$d_t - p_t = \frac{-\kappa}{1 - b} + \sum_{j=1}^{\infty} b^{j-1} E_t(r_{t+j}) - \sum_{j=1}^{\infty} b^{j-1} E_t(\Delta d_{t+j}). \quad (6)$$

As in Cochrane (2008), equation (6) can apply to real returns and real dividend growth by relating real return  $r_{t+1} - \pi_{t+1}$  to dividend growth less the inflation rate,  $\Delta d_{t+1} - \pi_{t+1}$ . So can it to excess returns by relating excess return  $r_{t+1} - r_t^f$  to dividend growth less the risk-free rate,  $\Delta d_{t+1} - r_t^f$ . Equation (6) shows that changes in the expected future dividend growth are an important source for movements in stock prices.

We assume that the log dividend process is the sum of random walk with drift

---

<sup>3</sup>We rule out the possibility of rational bubbles following the convention in the literature [see, for example, Cochrane (2008)].

and stationary components,

$$\begin{aligned} d_t &= w_{1,t} + w_{2,t}, \\ w_{1,t} &= (1 - \phi)\bar{w}_1 + \phi w_{1,t-1} + \eta_{1,t}, \\ w_{2,t} &= g + w_{2,t-1} + \eta_{2,t}, \end{aligned} \tag{7}$$

where both  $\eta_{1,t}$  and  $\eta_{2,t}$  are i.i.d. zero mean random variables with variance  $\sigma_1^2$  and  $\sigma_2^2$ , respectively, and  $g$  is a drift which is equal to the expected dividend growth rate when  $\phi = 1$ . We further assume that the persistency parameter  $\phi$  in the dividend process obeys

$$\phi = 1 - \frac{c}{T}, \tag{8}$$

where  $c$  is a positive number and  $T$  is sample size. This local-to-unity assumption has been used in many econometric works on predictive regressions following Phillips (1987) to describe nonstationary predictors with the aim of producing smooth asymptotics towards the unit root limit case as  $T \rightarrow \infty$ , see e.g. Elliot and Stock (1994) and Cavanagh, Elliot, and Stock (1995). In the present paper, we use this assumption for the dividend process so that the derived dividend-price ratio will preserve the same property.

Under the null hypothesis in which the expected excess returns are constant, the log of the dividend-price ratio is derived from equations (6)-(7) as

$$d_t - p_t = cons + \frac{1 - \phi}{1 - b\phi} w_{1,t}, \tag{9}$$

where  $cons = \frac{r-g-\kappa}{1-b} - \frac{(1-\phi)\bar{w}_1}{1-b\phi}$ . Note that the persistence of the log dividend-price ratio is governed by the parameter  $\phi$  in  $w_{1,t}$ .<sup>4</sup> That is, equation (9) links the persistence in the dividend process to that in the dividend-price ratio in predictive regression (1) where  $d_{t-1} - p_{t-1}$  is the regressor ( $x_{t-1}$ ). On the one hand, the dividend-price ratio is highly persistent but stable in finite samples under (8), which is compatible with empirical evidence and economic theory. On the other hand, it becomes constant at the limit, which is also compatible with economic theory: from equations (8)-(9) and using the definition of  $b$ , we have

$$\lim_{t \rightarrow \infty} \frac{D_t}{P_t} = \frac{R - G}{G}.$$

---

<sup>4</sup>We could generalize (8) to cover higher order AR models for the dividend-price ratio where the largest root is local-to-unity, and the other ones bounded away from the unit circle. However, there would be no further insights, while this modification affects only the long-run variance of the innovations  $v_t$ .

This is the well-known formula from the Gordon growth model for the price of a stock with the constant discount factor and the constant dividend growth rate. Equations (7)-(8) on the dividend process ensure the dividend growth to be stationary. In turn, identity (3) implies that the dividend-price ratio is stable because both the dividend growth and the gross return are stable, which enables us to perform the log linear approximation on equation (4) around the point  $\bar{P}/\bar{D} = G/(R - G)$ . In this sense, the log-linearized present value relation (6) can be interpreted as a dynamic generalization of the constant dividend-price ratio in the Gordon model.

The log stock return is derived from equations (5), (7), and (9),

$$r_t = \alpha + \frac{1-b}{1-b\phi} \left( \eta_{1,t} + \frac{1-b\phi}{1-b} \eta_{2,t} \right), \quad (10)$$

where  $\alpha = r + b(1-\phi)^2 \bar{w}_1 / (1-b\phi)$  represents a collection of constant terms. In the predictive regression,  $r_t$  is the dependent variable ( $y_t$ ). Under the null hypothesis that the expected excess returns are constant ( $\beta_0 = 0$ ),  $y_t$  is a prediction error ( $u_t$ ) plus constant. Accordingly, the error terms  $u_t$  and  $v_t$  are defined by

$$\begin{aligned} u_t &= \frac{1-b}{1-b\phi} \left( \eta_{1,t} + \frac{1-b\phi}{1-b} \eta_{2,t} \right) \\ v_t &= \frac{1-\phi}{1-b\phi} \eta_{1,t}. \end{aligned} \quad (11)$$

Then, the covariance matrix  $\Sigma = \Sigma(\phi, b)$  of  $u_t$  and  $v_t$  in predictive regression (1) has elements given by

$$\begin{aligned} \sigma_{uv}(\phi, b) &= \left( \frac{1-b}{1-b\phi} \right)^2 \left( \frac{1-\phi}{1-b} \right) \sigma_1^2, \\ \sigma_u^2(\phi, b) &= \left( \frac{1-b}{1-b\phi} \right)^2 \left( \sigma_1^2 + \left( \frac{1-b\phi}{1-b} \right)^2 \sigma_2^2 \right), \\ \sigma_v^2(\phi, b) &= \left( \frac{1-\phi}{1-b\phi} \right)^2 \sigma_1^2, \end{aligned} \quad (12)$$

where the covariance matrix is a function of the two economic parameters, the persistence parameter in the dividend process and the discount factor. Equation (12) explicitly shows that the innovations covariance matrix is not independent of the local-to-unity assumption on  $\phi$ . Torous, Valkanov, and Yan (2004), Moon, Rubia, and Valkanov (2004) and Gospodinov (2009) also make local-to-zero assumptions on  $\sigma_v^2$  to match stylized facts from data, but this is exogenously imposed and not driven by the economic model or primitive assumptions on other parameters such as  $\phi$ .



We have derived the covariance matrix of innovations distributions implied by the present value model under the null hypothesis and illustrated that the covariance matrix cannot be treated independently of the local to unity assumption on the regressor persistence. To analyze the power properties of regression based tests with this assumption, we also need to model (local) alternatives implied by the present value model.<sup>5</sup> For this, let us assume a simple local alternative,

$$E_t(r_{t+1}) = \beta^a(d_t - p_t), \quad (13)$$

where  $\beta^a$  is a non zero number. We will provide a specific assumption on  $\beta^a$  which depends on the sample size  $T$  in next section. For notational simplicity, we will use the same notation for relevant variables both under the null and alternative hypotheses. Then, as before, taking conditional expectation on return identity (5), iterating forward it, and assuming that no rational bubbles exist, we have the following present value identity under the alternative (13),

$$d_t - p_t = \frac{-\kappa}{1 - b - \beta^a} - \frac{1}{1 - \beta^a} \sum_{j=1}^{\infty} \left( \frac{b}{1 - \beta^a} \right)^{j-1} E_t(\Delta d_{t+j}), \quad (14)$$

where  $\beta^a$  is assumed to be small enough so that  $b/(1 - \beta^a) < 1$  and  $1 - b - \beta^a > 0$ .

Assuming the same dividend process as before, the log of the dividend-price ratio is derived from equations (7) and (14),

$$d_t - p_t = cons1 + \frac{1 - \phi}{1 - b\phi - \beta^a} w_{1,t}, \quad (15)$$

where *cons1* is a collection of constant terms. Finally, using  $d_t - p_t$  in (15) and  $\Delta d_{t+1}$  in (7), we obtain  $r_{t+1}$  in the return identity as well as  $u_{t+1}$  since  $r_{t+1} = E_t(r_{t+1}) + u_{t+1}$  where  $u_{t+1}$  is a collection of unpredictable components in the excess return. Then, in the predictive regression  $y_t$  is the sum of the expected excess return ( $E_{t-1}(r_t)$ ), a prediction error ( $u_t$ ), and a constant. Accordingly, the error terms under the local alternative are defined by

$$\begin{aligned} u_t &= \frac{1 - b - \beta^a}{1 - b\phi - \beta^a} \left( \eta_{1,t} + \frac{1 - b\phi - \beta^a}{1 - b - \beta^a} \eta_{2,t} \right) \\ v_t &= \frac{1 - \phi}{1 - b\phi - \beta^a} \eta_{1,t}, \end{aligned} \quad (16)$$

---

<sup>5</sup>We thanks Associate Editor to point out this aspect.

where  $1 - b\phi - \beta^a > 0$ . Finally, the covariance matrix  $\Sigma = \Sigma(\phi, b, \beta^a)$  of  $u_t$  and  $v_t$  in the predictive regression (1) has elements given by

$$\begin{aligned}\sigma_{uv}(\phi, b, \beta^a) &= \left( \frac{1 - b - \beta^a}{1 - b\phi - \beta^a} \right) \left( \frac{1 - \phi}{1 - b\phi - \beta^a} \right) \sigma_1^2, \\ \sigma_u^2(\phi, b, \beta^a) &= \left( \frac{1 - b - \beta^a}{1 - b\phi - \beta^a} \right)^2 \left( \sigma_1^2 + \left( \frac{1 - b\phi - \beta^a}{1 - b - \beta^a} \right)^2 \sigma_2^2 \right), \\ \sigma_v^2(\phi, b, \beta^a) &= \left( \frac{1 - \phi}{1 - b\phi - \beta^a} \right)^2 \sigma_1^2.\end{aligned}\tag{17}$$

We will use this covariance matrix implied by the present value model to study the power properties of regression-based tests in Section 3.

### 2.3 Discussion

The key assumption we made for deriving the covariance matrix (12) from the present value model of stock prices is that the dividend growth is predictable, in particular, based on its own past values. Specifically, we assumed that the log dividend follows a nonstationary process (7) in which a part of the dividend shock is permanent and the other is temporary. To test whether or not this assumption is reasonable, we have conducted variance ratio tests which are appropriate for our purpose since they provide valid information for identifying a specific economic model for which the temporary part of the shock implies predictability (negative autocorrelation) of the dividend growth.<sup>6</sup> In general, we find predictability of dividend growth over long horizons: the autocorrelations of dividend growth are negative over the all horizons considered, supporting assumption (7). For the sake of simplicity we relegate the details of our test procedures and results to Table 1 as well as to our working paper version.

We now discuss the implication of the local-to-unity assumption on the dividend process in equation (8) for statistical inference in the predictive regression and relate it to relevant studies in the literature.

We first heuristically show how the noise-to-signal ratio,  $\sigma_u/\sigma_v$ , implied by the present value model, affects statistical inference, paying a particular attention on its relation to the local-to-unity assumption on  $\phi$ .<sup>7</sup> For example,  $\sigma_u/\sigma_v$  mainly

---

<sup>6</sup>Complement to this test, we have also conducted a regression-based long-horizon test following Fama and French (1988). But the results remain unchanged and are reported in the working paper version of the article.

<sup>7</sup>The scaled correlation  $\sigma_{uv}/\sigma_v^2$ , which is also implied by the present value model, affects the

determines the variance of the estimated slope coefficient,  $\widehat{\beta}$ , and thus the rate of convergence to the true parameter value. Specifically, the ratio becomes infinity at the limit under the assumption (8) as can be easily seen in equations (12) and (17). Although we provide the formal analysis on the power properties of the regression tests based on the local-to-unity regressor in the next section, the key idea can be best illustrated writing from equations (9) and (11) (or analogously from equations (15) and (16) replacing  $1 - b$  by  $1 - b - \beta^a$ )

$$x_{t-1} \approx \frac{1 - \phi}{1 - b} w_{1,t-1}, \quad u_t \approx \eta_{1,t} + \eta_{2,t}, \quad (18)$$

where we ignore constant terms. The approximations in (18) suggest that the variance of the regressor may converge to zero rather than diverge to infinity despite of its non-stationary character. This is because  $(1 - \phi)$  can collapse to zero at a faster rate than the one at which the standard deviation of the near unit-root process  $w_{1,t-1}$  diverges with sample size. Note that in the predictive regression with constant  $\Sigma$ , an increment in  $\phi$  affects not only the skewness of the distribution but also the convergence rate of the estimated slope coefficient. However, it ignores the possibility that the scale factor  $(1 - \phi)$  also affects the regressor scaling and the convergence rate of the estimate. That is, larger values of  $\sigma_u/\sigma_v$  implied by the near unit value of  $\phi$  make the distribution of  $\widehat{\beta}$  wider and thus the estimated slope coefficient less informative.<sup>8</sup> On the other hand, an exogenous local-to-zero assumption on  $\sigma_v^2$  can make the convergence rate of  $\widehat{\beta}$  arbitrary.

The idea of investigating dependency between an economic parameter and the innovation covariance matrix is first proposed by West (2008) who introduced a local-to-unity assumption on the discount factor in the present value model of foreign exchange rates,

$$b = 1 - \frac{\delta}{\sqrt{T}}, \quad (19)$$

where  $\delta$  is a constant. He showed that the  $t$ -test is not consistent as the discount factor goes to one, while assuming the stationarity of the regressor. Combining

---

magnitude of the finite sample bias, while the relative quantity,  $\frac{\sigma_{uv}}{\sigma_v^2} / \frac{\sigma_u}{\sigma_v}$ , which is the contemporaneous correlation, affects the possible over-rejections of the  $t$ -test along with the persistent regressor.

<sup>8</sup>A similar idea can be applied to the cases in which  $\phi$  is strictly less than one. For example, the power of the  $t$ -test will be lower for higher values of  $\phi$  since  $\sigma_u/\sigma_v$  is larger. This is consistent with Cochrane (2008) who also reached the same conclusion that the power of the regression test would be lower as  $\phi$  increases, but based on a different approach.

his assumption with the assumption (8) would not change our conclusions as long as  $\phi$  converges to unity faster than the discount factor. Further, his assumption is valid in our model only for a particular solution that implies a zero value of the dividend-price ratio at the limit. That is, it only holds for the case where  $R = G$  at the limit, as can be easily checked from equation (3), which is a particular case of the Gordon model. Nevertheless, equation (18) implies that higher values of  $b$  may offset some power losses due to assumption (8) in finite samples.<sup>9</sup>

Finally, we relate the predictive regression in the present value model of stock prices to Cochrane (2008), who quantifies the relation between the predictability of stock returns and of dividend growth by deriving the following identity

$$\beta = 1 - b\phi + \beta_d, \quad (20)$$

from the bivariate regression (1), the regression of the log dividend growth on the log dividend-price ratio

$$\Delta d_t = \alpha_d + \beta_d(d_{t-1} - p_{t-1}) + u_t^d, \quad (21)$$

and identity (5). As Cochrane (2008) notes, the identity (20) implies that the dividend growth must be predictable and the slope coefficient  $\beta_d$  must be negative if the stock returns are unpredictable ( $\beta = 0$ ). He argues that joint hypothesis tests may yield more power under the assumption that  $\phi$  is strictly less than one.<sup>10</sup> Our present value model also satisfies this identity and implies  $\beta_d = b\phi - 1$  and  $u_t^d = \eta_{1,t} + \eta_{2,t}$  under the null hypothesis  $\beta = 0$  as can be easily checked from equations (7) and (9). However, assumption (8) along with higher values of the discount factor implies that  $\beta_d$  can be close to zero, even though  $\beta = 0$ . That is, the absolute value of  $\beta_d$  becomes smaller as  $\phi$  goes to one. At the same time, assumption (8) causes the same problem in the regression (21) as in the predictive regression (1) because  $sd(u_t^d)/\sigma_v = (\frac{1-b\phi}{1-\phi})\sqrt{(\sigma_1^2 + \sigma_2^2)/\sigma_1^2}$  also diverges to infinity under (8), implying that  $\hat{\beta}_d$  may not be informative and tests based on this regression may suffer from lack of power. This suggests an explanation for the difficulties of previous studies to find predictability in the dividend growth using

---

<sup>9</sup>In a related study, Cochrane (2008) also show that larger values of the discount factor would improve the power of the predictability tests, in particular for testing the predictability of long-horizon stock returns.

<sup>10</sup>For example, Engsted and Pedersen (2010) follow Cochrane (2008) and find the predictability of the dividend growth for international data.

a price-dividend ratio (see, for example, Cochrane (2008) and references therein). This result can also be compared to Engel and West (2005) who analytically show that fundamental variables are not helpful for predicting changes in exchange rates if they are generated from the typical present value model with near unit discount factor and nonstationary fundamentals.

### 3 The Power of Regression-Based Tests in the Present Value Model

It is now well established that local-to-unity assumptions on the roots of the regressor autoregressive polynomial affect the convergence rate of estimates in predictive regressions. These assumptions, however, have not been considered in an econometric framework where  $\Sigma$  is a function of  $\phi$  and thus of sample size. In this section, we study the power properties of regression-based tests for the predictability of stock returns generated from the present value model which incorporates the assumption of the local-to-unity in the regressor. Those tests include the conventional  $t$ -test as well as tests that are robust to the persistence and endogeneity problems such as CY's  $Q$ -test and JM's conditional test.

#### 3.1 Asymptotic Properties of Regression-Based Tests for Local-to-Unit-Root Processes

The null asymptotic distribution of the usual  $t$ -statistic for testing  $H_0 : \beta = \beta_0$  based on the OLS estimate  $\hat{\beta}$  of the predictive regression (1),

$$t(\beta_0) = \frac{\sum_{t=1}^T (x_{t-1} - \bar{x})(y_t - \beta_0 x_{t-1})}{\hat{\sigma}_u \left( \sum_{t=1}^T (x_{t-1} - \bar{x})^2 \right)^{1/2}},$$

is not standard under the local-to-unity assumption (8) with fixed  $\Sigma$ . Here  $\bar{x}$  is the sample mean of  $x_t$  and  $\hat{\sigma}_u^2$  is the residual variance. This distribution has two independent components, one depending on a standard normal random variable  $Z$  and the other one depending on a functional of a Ornstein-Uhlenbeck process,

$$t(\beta_0) \sim_a (1 - \rho_0^2)^{1/2} Z + \rho_0 \frac{\tau_c}{\kappa_c}, \quad (22)$$

see Elliot and Stock (1994), where  $\kappa_c = \left( \int \bar{J}_c^2(s) ds \right)^{1/2}$ ,  $\tau_c = \int \bar{J}_c(s) dW_v$ , and  $\rho_0 = \sigma_{uv}/(\sigma_u \sigma_v)$  is the correlation coefficient between  $u_t$  and  $v_t$  computed from  $\Sigma$ . The diffusion process  $\bar{J}_c(s)$  is the demeaned version of the Ornstein-Uhlenbeck process  $J_c$  defined by  $dJ_c(s) = cJ_c(s)ds + dW_v(s)$  in  $[0, 1]$  and  $J_c(0) = 0$  in terms

of a standard Wiener process  $W_v$ . The variable  $Z$ , which is independent of the functionals of  $J_c$ , leads to standard asymptotics when there is no endogeneity between the error terms, that is when  $\rho_0 = 0$ . Equation (22) shows that the usual  $t$ -test becomes inappropriate if  $c$  is fixed because it is not asymptotically pivotal.

Robust tests to the persistence (and endogeneity) problem such as JM's conditional test and CY's  $Q$ -test correct these problems to conduct asymptotic inference in predictive regressions. However, we show that these methods are seriously affected by the diminishing signal-to-noise ratio implied by the near unit  $\phi$  in our framework, which leads to a slower convergence rate of predictive regressions estimates and reduces the class of alternative hypotheses that these tests can detect consistently.

The different robust tests mainly differ on how they deal with the unknown value of  $\phi$ , since no uniformly most powerful (UMP) test exists without further restrictions, see Stock and Watson (1996). Using the conditionality principle, JM argue that tests for  $H_0 : \beta = \beta_0$  should be based on the conditional limit distribution of

$$R_\beta = \frac{\sigma_{u,v}^{-1}\sigma_v^{-1}}{T} \sum_{t=1}^T (x_{t-1} - \bar{x}) (y_t - \beta_0 x_{t-1} - b_{uv} \Delta x_t)$$

given the sufficient statistic  $(R_\phi, R_{\beta\beta}, R_{\phi\phi})$ , where  $R_\phi = \sigma_v^{-2} T^{-1} \sum_{t=1}^T x_{t-1} \Delta x_t - \rho_0 R_\beta / \sqrt{1 - \rho_0^2}$ ,  $R_{\beta\beta} = \sigma_v^{-2} T^{-2} \sum_{t=1}^T (x_{t-1} - \bar{x})^2$ , and  $R_{\phi\phi} = \sigma_v^{-2} T^{-2} \sum_{t=1}^T x_{t-1}^2$ , since this distribution does not depend on  $\phi$ . Here  $b_{uv} = \sigma_{uv} / \sigma_v^2$ ,  $\sigma_{u,v}^2 = \sigma_u^2 (1 - \rho_0^2)$ . Then, for a given value of  $\rho_0$ , a one-sided test rejects when the value of  $R_\beta$  exceeds a critical value  $C_\alpha(R_\rho, R_{\beta\beta}, R_{\phi\phi}; \rho_0)$  from the previous conditional distribution that preserves the asymptotic conditional  $\alpha$ -similarity property of the test, where  $\alpha$  is the desired size of the test. However, there does not exist a closed expression for such distribution under (8) and numerical methods are required, *cf.* Polk, Thompson and Vuolteenaho (2006).

By contrast, CY derive the  $Q$ -test invoking the Neyman-Pearson Lemma under the assumption that  $\phi$  is known a priori. Then, the  $Q$ -statistic for testing  $H_0 : \beta = \beta_0$  is a modified  $t$ -statistic under knowledge of  $\phi$  defined by

$$Q(\beta_0, \phi) = \frac{\sum_{t=1}^T (x_{t-1} - \bar{x}) (y_t - \beta_0 x_{t-1} - b_{uv} (x_t - \phi x_{t-1}))}{\sigma_{u,v} \left( \sum_{t=1}^T (x_{t-1} - \bar{x})^2 \right)^{1/2}},$$

resulting from regressing  $y_t - b_{uv}(x_t - \phi x_{t-1})$  onto a constant and  $x_{t-1}$ , assuming that  $\Sigma$  is known. Relaxing this last assumption is possible by replacing  $\Sigma$  with consistent estimates. This guarantees that the error term of this transformed regression is asymptotically uncorrelated with the regressor and when  $\phi$  is known the  $Q$ -statistic is asymptotically standard normal. The statistic can also be written as

$$Q(\beta_0, \phi) = \frac{\widehat{\beta} - \beta_0 - b_{uv}(\widehat{\phi} - \phi)}{\sigma_{u.v} \left( \sum_{t=1}^T (x_{t-1} - \bar{x})^2 \right)^{-1/2}},$$

exhibiting how the  $Q$ -statistic corrects the bias and side effects from estimation of  $\phi$ . CY propose to use Bonferroni's method to construct confidence intervals for  $\phi$  in order to avoid the dependence of the unfeasible  $Q$  on its unknown value.

CY show that tests based on  $Q$  are UMP within the class of invariant tests conditional on the ancillary statistic  $R_{\beta\beta}$  under normality and  $\phi$  known, but this property does not necessarily extend to the feasible version. On the other hand, JM's test is UMP within the class of conditionally unbiased tests and the Gaussian asymptotic power envelope is attainable in the absence of normality, but it does not mean that it dominates any other test, including the  $Q$ -test. Our asymptotic analysis focuses on examining the  $t$ - and  $Q$ -tests, but a parallel study of the conditional test properties would lead to similar conclusions, cf. Lemmas 3 and 4 in JM.

In the next result we justify that both the  $t$ - and  $Q$ -tests keep the same asymptotic null distribution when we take into account the variance-covariance matrix (12) implied by our present value model and the local-to-unity assumption under  $\beta_0 = 0$ . The proofs of this and next proposition are contained in Appendix.

**Proposition 1.** Under  $H_0$ , (8) and (12), if  $(\eta_{1,t}, \eta_{2,t})'$  is an homoskedastic martingale difference process with finite fourth moments and  $E[x_0^2] < \infty$ , the  $t$ -test statistic  $t(0)$  behaves as (22) while the  $Q$ -statistic behaves as

$$Q(0, \phi) \sim_a Z.$$

To study the power properties of regression based tests when the elements of  $\Sigma$  in (17) depend on  $T$ , we consider as in JM the following class of  $T^{-1}$ -local to

zero alternatives in the parameter  $\beta$  indexed by  $\zeta$ ,

$$H_{A,T}(\zeta) : \beta = \beta_T^a(\zeta) := \frac{\zeta}{T} \left( \frac{\sigma_{u.v}^2}{\sigma_v^2} \right)^{1/2}, \quad (23)$$

where  $\sigma_{u.v}^2 = \sigma_u^2 - \frac{\sigma_{uv}^2}{\sigma_v^2}$  and the normalization in  $\beta_T^a(\zeta)$  provides neat asymptotic results in terms only of  $\rho_0 = \lim_{T \rightarrow \infty} \sigma_{uv}/(\sigma_u \sigma_v)$ . From equations (8) and (17), the local alternative in (23) can be rewritten as

$$\beta_T^a(\zeta) \approx \frac{\zeta}{c} (1 - b - \beta_T^a(\zeta)) \left( \frac{1 - \rho_0^2}{\rho_0^2} \right)^{1/2} = \frac{\zeta}{c} (1 - b - \beta_T^a(\zeta)) \frac{\sigma_2}{\sigma_1}, \quad (24)$$

because  $\rho_0^2 = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$  under both (12) and (17). We could solve equation (24) to find the limiting value of  $\beta_T^a(\zeta)$ , obtaining after simple algebra  $\beta^a = \zeta \sigma_2(1 - b)/(c\sigma_1 - \zeta \sigma_2)$ , which is arbitrarily small in absolute value if  $|\zeta|$  is chosen small enough.

In contrast to the case of the predictive regression with an exogenous fixed  $\Sigma$ , parameterized independently of the value of  $\phi$ , the parameter  $\beta^a$  values defined by equation (24) are no longer  $T^{-1}$  local to the null  $H_0 : \beta = 0$  (i.e.  $H_{A,T}(0)$ ) but bounded away from zero as  $T$  increases. This results because  $\sigma_v$  in equation (12) is converging to zero with  $T$  and thus  $\sigma_u^2/\sigma_v^2$  is approaching infinity asymptotically. This asymptotic behavior of the noise-to-signal ratio significantly affects the convergence rates of estimates in the predictive regression.

Under the covariance matrix (17) implied by our present-value model under a small degree of predictability, the OLS estimate  $\hat{\beta}$  is inconsistent for any true value of  $\beta$ , so that the corresponding sample  $R^2$  coefficient is converging to zero under  $H_{A,T}(\zeta)$ , since, after standardization by  $T$ ,  $R^2$  has the same bounded limit as the square of the corresponding  $t$ -statistic. This explains the lack of consistency of regression-based tests against  $H_{A,T}(\zeta)$  despite it specifies that  $\beta_T^a(\zeta)$  is away from zero even when  $T \rightarrow \infty$ . This is formally stated in the following result.

**Proposition 2.** Under  $H_{A,T}(\zeta)$ , (8), (17), and the regularity conditions of Proposition 1, the  $t$ -statistic behaves as

$$t(0) \sim_a \zeta (1 - \rho_0^2)^{1/2} \kappa_c + \rho_0 \frac{\tau_c}{\kappa_c} + (1 - \rho_0^2)^{1/2} Z,$$

while the  $Q$ -statistic behaves as

$$Q(0, \phi) \sim_a \zeta \kappa_c + Z.$$



Therefore, the  $Q$ -test has only nontrivial power against alternatives that are fixed with respect to the null  $\beta = 0$ , compared to the usual  $T$  rate of convergence of  $\hat{\beta}$  under (8) and fixed innovation covariance matrix. Similarly, JM's unbiased test conditional on ancillary statistics has only nontrivial power against fixed alternatives as given by (24), while, obviously, the  $t$ -statistic shows the same asymptotic behaviour as the  $Q$ -statistic when  $\rho_0 = 0$ . These results also imply that predictive regression-based tests are inconsistent when the data are generated from our present value model of stock prices since power does not increase with sample size. The results also go through in the exactly same way for the tests of dividend growth predictability by the dividend-price ratio in equation (21) in that the local alternatives yield in equation (24).

To provide further intuition on the power properties of the  $Q$ -test we can consider the non-standardized local alternatives

$$H_{A,T}^*(\mu) : \beta = \beta_T^{a*}(\mu) := \frac{\mu}{T},$$

which guarantee that (17) holds. Then, we find that under  $H_{A,T}^*(\mu)$

$$Q(0, \phi) \approx \mu \frac{1 - \phi}{1 - b} \left( \frac{\rho_0^2}{1 - \rho_0^2} \right)^{1/2} \kappa_c + Z, \quad (25)$$

up to  $O_p(T^{-1})$  terms, noting that  $\beta_T^{a*}(\mu) = O(T^{-1})$ . This shows that the power of the  $Q$ -test, driven by the first factor on the right hand side of equation (25), increases with  $\rho_0^2$ , but it is only trivial when  $\phi \rightarrow 1$ . A similar analysis is possible for the  $t$ -test, leading to similar qualitative conclusions, though its drift depends only linearly on  $|\rho_0|$ . However, if  $1 - b$  is small relative to  $1 - \phi$ , the power of the  $Q$ -test may not be negligible asymptotically, as happens under the assumption that the discount factor tends to be one with sample size in equation (19). We further explore these issues in our simulation analysis in Section 4.

### 3.2 Discussion

We have shown that the regression-based tests may suffer from lack of power under (8) once the scaling effects of a local-to-unity  $\phi$  on the covariance matrix of innovations and on the convergence rate of the estimate are taken into account. In this subsection, we consider some alternative tests which could overcome this

problem. Nonparametric orthogonality tests are one possible candidate: for example, the sign tests of Campbell and Dufour (1995) and the covariance-based orthogonality tests of Maynard and Shimotsu (2009). Like the robust tests, those tests allow for the possible feedback from innovations to future values of the regressors as well as strong persistence in the regressors. However, we find that those nonparametric tests are also inconsistent and thus exhibit power deficiency under (8) as typical regression tests.

To see this, first note that the sign tests of Campbell and Dufour (1995) have pivotal finite sample distributions and standard asymptotics with usual  $T^{1/2}$  rate of convergence under the null assuming a mediangale type of condition for the error term, irrespective of the dynamic properties of the regressor. Similar properties can be achieved in the case of estimation of the centering parameter. However, under (8), the sign test only has power against fixed alternatives and cannot detect local alternatives converging to the null with sample size, implying that the test is inconsistent as the previous regression-based tests. The reason for this asymptotic behaviour is that the random error term  $x_{t-1}u_t$  dominates the sign of  $(y_t - \alpha)x_{t-1} = \beta^a x_{t-1}^2 + x_{t-1}u_t$  under the alternative, because  $\beta^a x_{t-1}^2$  is of smaller order of magnitude, cf. (13), even for fixed  $\beta^a$ , explaining the low power of sign tests in applications compared with other methods [see, for instance, the conclusions in Maynard (2006)].

A similar effect arises with the covariance-based orthogonality tests of Maynard and Shimotsu (2009). The covariance estimate they propose is only  $(T/m)^{1/2}$ -consistent with local-to-unity regressors and standard fixed covariance assumptions, where  $m$  is an increasing bandwidth parameter. However, under  $H_{A,T}(\zeta)$  and (12), the drift of their  $t$ -statistic is only of order  $O_P((m/T)^{1/2}) = o_P(1)$ , so its asymptotic power against fixed alternatives implied by equation (24) is just equal to the nominal size.

In general, we can conclude that existing regression-based predictive tests have power problems when applied to present value models with local-to-unity regressors. These models suggest that there should be only very weak (and shrinking) predictive content in the dividend-price ratio when its root is close to unity. For this reason, even a test with good power under standard conditions may not be able to detect such persistent regressors. Only under specific alternatives that

imply observable predictors with enough signal-to-noise ratio it could be possible that standard methods can lead to consistent procedures.

## 4 Monte Carlo Experiments

We use two models for Monte Carlo experiments. One derives the covariance matrix from the present value model of stock prices, which is the main focus of the present paper; the other assumes a constant covariance matrix for comparison. For each model, we calculate the size and power of three regression-based tests which include the conditional test, the  $Q$ -test, and the conventional  $t$ -test.<sup>11</sup>

We use equations (7)-(12) as a data generating process for the present value model. We set  $\sigma_1/\sigma_2 = 2$  and  $b = 0.9$ . We vary the values of  $c$  in equation (8) from 20 to 1. Then, for example, the implied values of  $\phi$  are 0.990, 0.996, 0.998 and 0.999 for sample sizes of  $T = 100, 250, 400$  and 1000, respectively, when  $c = 1$ . For comparison, we also use the same equations as a data generating process for the model with a constant covariance matrix. But, we replace  $\phi$  in equation (12) with  $\bar{\phi}$  so that the covariance matrix does not change with  $T$  or  $c$ . We set  $\bar{\phi} = 0.9$  which is equivalent to the value of  $\phi$  when  $T = 100$  and  $c = 10$  in the present value model. With these parameter values, the contemporaneous correlation  $\rho$  is about 0.73 in the model with the constant covariance matrix. On the other hand, the correlation is endogenously determined in the present value model and changes with  $T$  and  $c$ : it increases with  $\phi$ .

### 4.1 The Size

Table 2 reports the size of the three regression tests using 1,000 repetitions. All tests are conducted for the conventional significant levels against right-tail, left-tail and two-sided alternatives. To conserve space, we only report the results of the tests conducted for the 5% significant level against left- and right-tail alternatives. Overall we find that each of the three tests produces similar size in both models, consistent with the analytical results in the previous section (see Proposition 1). In terms of generating the right size, we find that the conditional test performs best among the tests considered. It produces rejection rates close to the nominal

---

<sup>11</sup>We obtain the algorithms for the conditional test from Polk's homepage ([http : //personal.lse.ac.uk/POLK/research/work.htm](http://personal.lse.ac.uk/POLK/research/work.htm)) and for the  $Q$ -test from Yogo's homepage ([http : //finance.wharton.upenn.edu/~yogo/](http://finance.wharton.upenn.edu/~yogo/)).

value, regardless of the presence of high contemporaneous correlation between  $u_t$  and  $v_t$  and of strong persistence in the regressor.

We now discuss in detail the results from each of the three tests, respectively. We find that the conventional  $t$ -test produces significant size distortions, confirming the results from previous studies [see, e.g., Mankiw and Sharpiro (1986), CY, and JM]. This result is expected because our parameterization sets very high values of  $\phi$  and generates high values of  $\rho$ . For example, the empirical sizes of the 5% left-tail  $t$ -test are 32.3, 34.6, 33.4 and 31.2% for  $T = 100, 250, 400$  and 1000, respectively, in the present value model with  $c = 1$  where the associated contemporaneous correlations are 0.88, 0.89, 0.89 and 0.89. In the model with the constant covariance matrix, the rejection rates of the same test are 25.3, 25.8, 26.7, and 24.3% for the same values of  $T$  and  $c$ . The main difference in the rejection rates between the two models is from the different degree of the contemporaneous correlations. As Mankiw and Sharpiro (1986) illustrate, the overrejection rates of the conventional  $t$ -test increase with higher contemporaneous correlations for a given persistent level.

JM's conditional test generates rejection rates very close to the nominal value in both tails. The results are similar in both models considered. Neither are the properties of the conditional test sensitive to small sample sizes. In this sense, the test may be best suited for studying the predictability of asset returns when there is doubt on the validity of statistical inference due to bias distortions linked to the strong persistence in the regressor. On the other hand, the  $Q$ -test produces size distortions in some cases in that it tends to be over-sized in one-tail and under-sized in the other-tail.<sup>12</sup> For example, the empirical sizes of the 5% right-tail  $Q$ -test are 0.6, 0.6, 1.0 and 4.2% and those of the 5% left-tail  $Q$ -test are 7.6, 7.3, 6.3, and 2.9% for  $T = 100, 250, 400$ , and 1000, respectively, in the present model with  $c = 1$ . This rejection pattern also appears to be very similar in the model with the constant covariance matrix.

## 4.2 The Power

In order to investigate the power properties of the regression tests, we consider the following local alternative:  $\beta = \mu/T$ . We use two different values of  $\mu$ :  $\mu = -10$

---

<sup>12</sup>JM also show that the  $Q$ -test has size distortions in some models [see JM (p. 701)].

and  $\mu = 10$ . The values of all other parameters are the same as in the previous subsection.

Table 3 reports the power of the three regression tests against the above local alternative. We find that the results from Monte Carlo experiments confirm the prediction of our asymptotic analysis. All the tests are not consistent in the present value model: the power of the regression tests decreases with sample size  $T$  for a given  $c$ . Further it increases with  $c$  in the present value model which can be confirmed easily in equation (24). This is true even if size-corrected power is analyzed.<sup>13</sup> On the other hand, the power of the regression tests appears to be stable with sample size for a given  $c$  in the constant covariance matrix model. Below, we mainly discuss the power properties of the conditional test between the two models since the other two tests also exhibit similar properties.

In the present value model the rejection rates of the 5% right-tail conditional test are 23.5, 12.2, 9.8 and 7.2% for  $T = 100, 250, 400$  and 1000, respectively, when  $c = 5$  and  $\mu = 10$ , clearly suggesting that the conditional test is not consistent. On the other hand, for the same parameter values of  $c$  and  $\mu$ , the rejection rates of the same test are 45.2, 46.9, 46.5 and 48.9% in the model with the constant covariance matrix, indicating that rejection rates do not vary over sample size as predicted by standard results. These rejection rates are in sharp contrast with those from the present value model. Further, for a given sample size of  $T = 100$  and  $\mu = 10$ , the rejection rates of the same test are 51.8, 36.5, 23.5 and 9.6% with  $c = 20, 10, 5$  and 1 in the present value model. On the other hand, for the same values of  $T$  and  $\mu$ , the rejection rates are 26.6, 36.5, 45.2 and 54.9% with  $c = 20, 10, 5$  and 1 in the constant covariance model, indicating that the power pattern of the conditional test with  $c$  in the present value is the opposite to that in the constant covariance model, holding everything else constant. These findings agree with the intuition that power increases with the regressor persistence (i.e. with lower  $c$ ) if the noise-to-signal ratio is fixed, but in the present value model this effect is dominated by the scaling effect on innovations of the dividend-price ratio, so a lower  $c$  implies reduced power, in agreement with equation (25) which shows a drift of the test statistics proportional to  $1 - \phi = c/T$ .

---

<sup>13</sup>For the sake of simplicity, we do not present the results of the size-corrected power but they are available upon request.

## 5 Conclusions

This paper investigates, both analytically and by simulation, the power properties of regression-based tests when the assumption of the local-to-unity regressor persistence is linked to the determination of the covariance matrix of innovations in the bivariate predictive regressions. We show the possibility that regression-based tests including robust tests to the persistent problem are inconsistent and thus have power deficiency if the time series are generated from the typical present value models. The main reason is that the regressor persistence parameter affects the covariance matrix of stochastic disturbance terms and thus the convergence rate of the estimated slope coefficient in those models. This potential effect on the power properties has not been taken into account in the previous studies mainly because the assumptions on the regressor persistence and the covariance matrix were treated independently.

Despite of the implication of power deficiency, our framework, in which the covariance of innovations is a function of the regressor persistence, overcomes the counterintuitive implication of returns being dominated by a nearly nonstationary component under the alternative hypothesis in the predictive regression because the near unit root component has a small (shrinking) variance.<sup>14</sup> Interestingly, alternatives under which returns are predictable but stationary lead to a similar conclusion to our framework in terms of poor power. In this sense, our results support the intuition that a mismatch in the persistence of the regressor and of the dependent variable is typically associated with low power predictive tests.

---

<sup>14</sup>We thanks an anonymous referee to point out this aspect.

## Appendix: Proofs

**Proof of Proposition 1.** The proof follows similarly to Appendix B in CY. First, we can write the  $t$ -statistic under  $H_0$  as

$$t(0) = \frac{T^{-1} \sum_{t=1}^T x_{t-1}^\mu u_t}{\hat{\sigma}_u \left( T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2} \right)^{1/2}} \quad (26)$$

where  $x_{t-1}^\mu$  is the demeaned  $x_{t-1}$  data. Now we apply the usual convergence results of sample moments of  $(\eta_{1,t}, \eta_{2,t})'$  and their partial sums, see for instance Lemma A.1 in CY, to a standardized version of the  $x_t$  series to control the degeneracy of  $\sigma_v$  to zero with  $T$ , which otherwise satisfies usual regularity conditions. Thus the scaled sum

$$T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2} / \sigma_v^2 = T^{-2} \sum_{t=1}^T w_{1,t-1}^{\mu 2} / \sigma_1^2 \quad (27)$$

converges to  $\kappa_c^2$  from (9) and (12). Next, we use the decomposition of  $u_t$  into  $v_t$  and an orthogonal component  $s_t$  with variance one,

$$\frac{u_t}{\sigma_u} = \rho \frac{v_t}{\sigma_v} + (1 - \rho^2)^{1/2} s_t,$$

where the convergence of

$$\frac{\rho}{T\sigma_v} \sum_{t=1}^T x_{t-1}^\mu \frac{v_t}{\sigma_v} = \frac{\rho}{T} \sum_{t=1}^T w_{1,t-1}^\mu \eta_{1,t} / \sigma_1^2 \quad (28)$$

to  $\rho_0 \tau_c$ , as well as that of

$$\frac{T^{-1} \sum_{t=1}^T x_{t-1}^\mu s_t}{\left( T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2} \right)^{1/2}}$$

to a standard normal, follow by the same normalizing argument and the uncorrelation of  $s_t$  and  $x_t$ . Then we use the fact that the residual variance  $\hat{\sigma}_u^2$  converges to  $\sigma_u^2$  because the regression coefficient is asymptotically bounded but  $x_t$  degenerates to zero.

The properties of the unfeasible  $Q$ -statistic  $Q(0, \phi)$  are derived similarly, noting that the contribution from (28) is canceled by the correction term and the normalization by  $(1 - \rho^2)^{1/2} \rightarrow (1 - \rho_0^2)^{1/2}$ .  $\square$

**Proof of Proposition 2.** The proof follows similarly to that in Proposition 1

using equation (24) to write the  $t$ -statistic under  $H_{1,T}(\zeta)$  and (17) as

$$t(0) \rightarrow_p T \frac{\zeta}{c} (1-b-\beta^a) \left( \frac{1-\rho_0^2}{\rho_0^2} \right)^{1/2} \frac{\left( T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2} \right)^{1/2}}{\widehat{\sigma}_u} + \frac{T^{-1} \sum_{t=1}^T x_{t-1}^\mu u_t}{\widehat{\sigma}_u \left( T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2} \right)^{1/2}}. \quad (29)$$

Now we have that

$$\frac{T\sigma_v}{\widehat{\sigma}_u} \frac{\zeta}{c} (1-b-\beta^a) \left( \frac{1-\rho_0^2}{\rho_0^2} \right)^{1/2} \rightarrow_p \frac{c\sigma_1}{(\sigma_1^2 + \sigma_2^2)^{1/2}} \frac{\zeta}{c} \left( \frac{1-\rho_0^2}{\rho_0^2} \right)^{1/2} = \zeta (1-\rho_0^2)^{1/2}$$

from the definition of  $\rho_0^2$  and that still  $\widehat{\sigma}_u \rightarrow_p \sigma_u \sim (\sigma_1^2 + \sigma_2^2)^{1/2}$  as  $T \rightarrow \infty$  under (17). Then the convergence of the first term in (29) follows from (27) and that of the second one from the proof of Proposition 1.

The analysis of the unfeasible  $Q$ -statistic  $Q(0, \phi)$  under  $H_{A,T}(\zeta)$  follows as in Proposition 1 and the normalization of the drift of  $t(0)$ .  $\square$



## References

- Campbell B, Dufour JM. 1995. Exact nonparametric orthogonality and random walk tests. *The Review of Economics and Statistics* 77 (1): 1-16.
- Campbell JY, Shiller RJ. 1987. Cointegration and tests of present value models. *Journal of Political Economy* 95 (5): 1062-1088.
- Campbell JY, Shiller RJ. 1988a. Stock prices, earnings, and expected dividends. *Journal of Finance* 43 (3): 661-676.
- Campbell JY, Shiller RJ. 1988b. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1 (3): 195-228.
- Campbell JY, Thompson SB. 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 21 (4): 1509-1531.
- Campbell JY, Yogo M. 2006. Efficient tests of stock return predictability. *Journal of Financial Economics* 81 (1): 27-60.
- Cavanagh, C., Elliott, G., Stock, J., 1995. Inference in models with nearly integrated regressors. *Econometric Theory* 11: 1131-1147.
- Cochrane, JH. 2008. The dog that did not bark: a defense of return predictability. *The Review of Financial Studies* 21 (4): 1533-75.
- Elliott G, Stock JH. 1994. Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory* 10: 672-700.
- Engel C. 1996. The Forward discount anomaly and the risk premium: a survey of recent evidence. *Journal of Empirical Finance* 3: 123-192.
- Engel C, West K. 2005. Exchange rates and fundamentals. *Journal of Political Economy* 113: 485-517.
- Engsted, T, Pedersen TQ. 2010. The dividend-price ratio does predict dividend growth: International evidence. *Journal of Empirical Finance* 17: 585-605.

- Fama E, French K. 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22 (1): 3-25.
- Gospodinov N, 2009. A New Look at the Forward Premium Puzzle. *Journal of Financial Econometrics*, 7(3): 312-338.
- Jansson M, Moreira HJ. 2006. Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74(3): 681-714.
- Lewis K. 1995. Puzzles in international financial markets. In *The Handbook of International Economics* 3, Grossman G, Rogoff K (eds). North Holland: Amsterdam.
- Mankiw G, Shapiro M. 1986. Do we reject too often? small sample properties of tests of rational expectations models. *Economics Letters* 20: 139-145.
- Maynard A, Shimotsu K. 2009. Covariance-based orthogonality tests for regressors with unknown persistence. *Econometric Theory* 25 (1): 63-116.
- Maynard, A. 2006. The forward premium anomaly: statistical artifact or economic puzzle? New evidence from robust tests. *Canadian Journal of Economics* 39 (4): 1244-1281.
- Moon S, Velasco C. 2010. Predictability tests for the expectations hypothesis in the presence of MA disturbances. Universidad Carlos III de Madrid, working paper.
- Moon, R., Rubia, A. and Valkanov, R. 2004. Long-Horizon Regressions when the Predictor is Slowly Varying, mimeo.
- Phillips, P. 1987. Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535-547.
- Polk C., Thompson S., Vuolteenaho T. 2006. Cross-sectional forecasts of the equity premium. *Journal of Financial Economics* 81 (1): 101-141.
- Poterba, J., and L. Summers 1988. Mean Reversion in Stock Prices: Evidence and Implications. *Journal of Financial Economics*, 22: 27-59.

- Stock, J. H. and M. W. Watson, 1996. Confidence Sets in Regressions with Highly Serially Correlated Regressors, Working Paper, Harvard University.
- Torous, W., R. Valkanov and S. Yan, 2004. On Predicting Stock Returns with Nearly Integrated Explanatory Variables, *The Journal of Business* 77(4), 937-966.
- Valkanov, R. 2003. Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics* 68, 201-32.
- West K. 2008. Econometric analysis of present value models when the discount factor is near one. University of Wisconsin, working paper.

**Table 1.** Predictability of dividend growth in US Stock Markets

Panel A. 1926-2008					Panel B. 1940-2008			
h	Nominal Dividend		Real Dividend		Nominal Dividend		Real Dividend	
	$\widehat{VR}(h)$	p-val	$\widehat{VR}(h)$	p-val	$\widehat{VR}(h)$	p-val	$\widehat{VR}(h)$	p-val
2	0.98	0.44	0.87	0.12	0.77	0.03	0.77	0.03
3	0.98	0.48	0.75	0.06	0.70	0.04	0.67	0.03
4	1.01	0.53	0.69	0.06	0.69	0.08	0.62	0.03
5	0.97	0.48	0.63	0.05	0.65	0.08	0.56	0.03
6	0.84	0.30	0.51	0.02	0.56	0.05	0.48	0.02
7	0.73	0.19	0.42	0.01	0.48	0.03	0.40	0.01
8	0.62	0.11	0.32	0.00	0.43	0.03	0.34	0.00
9	0.55	0.08	0.27	0.00	0.42	0.03	0.31	0.01
10	0.56	0.11	0.27	0.00	0.40	0.03	0.30	0.01
11	0.58	0.14	0.27	0.00	0.40	0.04	0.31	0.01
12	0.57	0.15	0.26	0.00	0.38	0.04	0.31	0.02
13	0.56	0.15	0.24	0.00	0.36	0.04	0.30	0.02
14	0.55	0.16	0.23	0.00	0.33	0.04	0.29	0.02
15	0.57	0.18	0.25	0.01	0.32	0.03	0.28	0.02
16	0.59	0.22	0.27	0.01	0.31	0.04	0.28	0.02

Note: This table reports the results from the long-horizon tests for the predictability of the dividend growth in U.S. stock markets using the variance ratio tests developed by Moon and Velasco (2010). Define the population variance ratio,  $VR(h)$ , exploiting that the variance of the sum of  $h$  consecutive dividend growths should be  $h$  times greater than that of  $\Delta d_t$  under the null hypothesis of no predictability of dividend growth,

$$VR(h) = \frac{Var(\sum_{i=0}^{h-1} \Delta d_{t+i})}{hVar(\Delta d_t)} = 1 + 2 \sum_{i=1}^{h-1} \left(1 - \frac{i}{h}\right) \gamma(i),$$

where  $h$  represents a holding period horizon and  $\gamma(i) = Cov(\Delta d_t, \Delta d_{t+i})/Var(\Delta d_t)$  denotes the autocorrelation of dividend growth between time  $t$  and  $t + i$ .  $VR(h)$  should be equal to one as long as dividend growth is not serially correlated. If the dividend growth is positively correlated,  $VR(h)$  should be greater than one; if the dividend growth is negatively correlated,  $VR(h)$  should be less than one.

The tests are conducted based on critical values from the empirical distribution generated by the wild bootstrap method following Moon and Velasco (2010). Annual dividends on the CRSP value-weighted index are used for our study. Following Cochrane (2008), dividend growth is calculated using the CRSP value-weighted indexes with/without dividend payments. Real dividends are constructed using the CPI data (the price series of “all urban consumers”) obtained from Bureau of Labor Statics.

Panel A reports the results from the sample period of 1926-2008, while Panel B reports the results from the sample period of 1940-2008.  $\widehat{VR}(h)$  is the estimate of variance ratios for the aggregation value  $h$  and ‘p-val’ is the p-value of  $\widehat{VR}(h)$ .

Data source: the Center for Research in Security Prices (CRSP).

**Table 2.** The Size of the Regression-Based Tests

$\mu$	c	b	T	Conditional test				$Q$ -test				$t$ -test			
				100	250	400	1000	100	250	400	1000	100	250	400	1000
Present value model															
0	20	0.9	5L	6.3	5.4	5.7	4.4	8.5	6.2	5.2	3.7	10.4	11.0	11.7	10.2
			5R	4.1	4.7	4.6	4.1	1.1	0.8	1.2	1.2	3.1	2.4	2.1	1.3
0	10	0.9	5L	5.5	5.6	5.1	4.3	7.0	5.5	5.3	4.8	13.3	14.1	14.5	13.3
			5R	4.6	3.7	4.4	4.7	0.7	1.2	1.1	1.3	1.7	1.0	1.1	0.5
0	5	0.9	5L	5.0	5.4	4.3	4.4	6.2	6.2	6.1	5.7	17.1	18.9	18.2	17.0
			5R	5.0	4.9	5.3	5.9	0.8	0.9	1.0	1.7	0.5	0.5	0.5	0.4
0	1	0.9	5L	4.8	5.2	4.6	5.1	7.6	7.3	6.3	2.9	32.3	34.6	33.4	31.2
			5R	5.4	5.8	5.3	6.6	0.6	0.6	1.0	4.2	0.1	0.0	0.3	0.1
Constant covariance matrix															
0	20	0.9	5L	6.2	5.5	6.0	5.4	9.4	6.2	5.6	4.6	10.8	10.7	11.1	9.3
			5R	4.4	4.7	4.8	4.5	0.7	0.9	1.3	1.1	2.6	2.3	2.1	1.8
0	10	0.9	5L	5.5	5.8	5.7	4.5	7.0	5.9	6.1	4.6	13.3	13.4	12.6	11.6
			5R	4.6	4.1	4.5	4.5	0.7	0.8	1.1	1.3	1.7	1.6	1.5	1.0
0	5	0.9	5L	5.3	5.9	5.0	4.2	6.0	6.4	5.6	5.7	16.2	17.0	15.8	14.9
			5R	5.1	4.7	4.9	4.8	0.8	1.0	1.1	1.0	1.0	0.9	0.9	1.1
0	1	0.9	5L	5.0	6.0	5.0	5.3	7.4	8.5	6.6	6.1	25.3	25.8	26.7	24.3
			5R	5.3	4.7	5.4	5.9	0.9	1.6	1.3	1.0	0.3	0.1	0.6	0.3

Note: We use equations (7)-(12) as a data generating processes. We consider four values of  $c$  which determines the value of  $\phi$  in equation (8). We set  $\sigma_1/\sigma_2 = 2$ . We conduct 1000 simulations and, for each simulation, generate four samples with the size of  $T = 100, 250, 400$ , and 1000, respectively. All tests are conducted for the 5% significant level against left-tail (5L) and right-tail (5R) alternatives.

**Table 3.** The Power of the Regression-Based Tests under the local alternative:  $\beta = \mu/T$ 

$\mu$	c	b	Conditional test				$Q$ -test				$t$ -test				
T			100	250	400	1000	100	250	400	1000	100	250	400	1000	
Present value model															
-10	20	0.9	5L	59.6	21.4	14.5	6.8	65.4	22.8	15.1	7.1	72.3	34.1	26.9	15.8
-10	10	0.9	5L	35.4	13.1	9.2	5.4	49.0	16.3	11.0	6.5	68.9	32.7	24.8	17.6
-10	5	0.9	5L	15.6	8.6	6.4	5.0	29.8	11.6	9.3	6.5	59.1	33.2	26.5	19.5
-10	1	0.9	5L	6.3	6.2	4.9	5.4	10.8	7.8	7.3	2.9	44.6	40.3	36.1	32.5
10	20	0.9	5R	51.8	23.0	14.6	9.2	25.2	9.3	5.3	2.5	43.6	15.0	8.6	3.0
10	10	0.9	5R	36.5	16.1	11.9	6.7	19.0	6.4	3.3	2.0	26.7	8.1	4.1	1.3
10	5	0.9	5R	23.5	12.2	9.8	7.2	12.2	3.5	2.7	2.1	13.7	2.5	1.8	0.6
10	1	0.9	5R	9.6	7.0	6.4	6.8	2.5	1.0	1.5	4.4	0.8	0.1	0.3	0.1
Constant covariance matrix															
-10	20	0.9	5L	29.3	25.7	27.2	27.4	35.7	27.7	28.0	24.9	45.4	40.5	43.0	42.6
-10	10	0.9	5L	35.4	33.5	34.4	32.4	49.0	40.1	43.4	38.6	68.9	62.3	65.2	62.0
-10	5	0.9	5L	42.8	40.4	40.9	39.3	67.4	61.8	62.9	60.6	87.6	85.6	86.3	86.0
-10	1	0.9	5L	59.6	61.7	62.3	59.6	84.8	82.0	85.8	84.0	97.5	97.8	97.7	97.8
10	20	0.9	5R	26.6	28.6	28.2	29.6	10.0	12.0	12.8	14.4	19.1	20.6	20.3	21.9
10	10	0.9	5R	36.5	39.6	38.4	39.9	19.0	22.1	22.7	25.4	26.7	28.9	27.9	28.9
10	5	0.9	5R	45.2	46.9	46.5	48.9	30.0	33.4	32.6	35.3	32.2	35.5	34.6	36.9
10	1	0.9	5R	54.9	53.5	54.4	56.2	45.5	48.8	48.6	52.0	42.1	40.6	41.6	41.6

Note: see note in Table 2.